知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum

E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【学术探索】

影响电影微博互动效果的隐藏主题探究方法及应用

张新香 赵彩霞
中南财经政法大学信息与安全工程学院 武汉 430073

摘要：[**目的** / **意义**] 探究影响电影微博互动效果的隐藏主题能发掘用户关注的热点问题，为企业提供有效的营销策略。[**方法** / **过程**] 从新浪微博上爬取 2017 年上映的 123 部电影的热门微博，采用主题建模方法挖掘电影微博文本中的隐藏主题，利用回归方法分析隐藏主题对电影微博互动效果的影响。[**结果** / **结论**] 结果发现 6 个可解释主题：电影人物、电影宣传、互动营销、电影内容、电影评价和线下活动，其中电影宣传、互动营销、电影内容和电影评价 4 个主题正向影响电影微博的互动效果；同时发现用户粉丝数和话题讨论热度正向影响电影微博的互动效果。

关键词：电影微博　互动效果　主题模型　LDA
分类号：G206
DOI：10.13266/j.issn.2095-5472.2020.027

引用格式：张新香 , 赵彩霞 . 影响电影微博互动效果的隐藏主题探究方法及应用 [J/OL]. 知识管理论坛 , 2020, 5(5): 283-291[引用日期]. http://www.kmf.ac.cn/p/223/.

微博作为新媒体时代的网络社交平台，因具有互动性、公开性、便捷性等优点 [1] 得到了千万网民的喜爱，发展迅速。同时微博平台因其短小而丰厚、平等而互动、快速而价优等优势 [2] 也得到众多企业青睐，其中就包括影视企业。《2018 年度微博电影白皮书》报告显示，2018 年 82 部票房过亿的电影，100% 开通了官方微博，全年累计电影阅读点评总量达 568 亿 [3]。事实数据说明微博平台拥有海量的用户，同样正成为影视企业进行产品营销的重要平台。

企业微博营销效果的一个重要衡量指标是微博的互动效果 [4]，即微博所引发用户的点赞、转发和评论的数量，故分析电影微博互动效果的影响因素是重要且有意义的。影响微博互动效果的因素有很多，其中，微博主题是一个较难获取和分析的因素，而用户往往对关注度高的主题有更强的互动意愿。鉴于此，本文尝试找出电影微博中的隐藏主题及其对互动效果影

基金项目：本文系国家社会科学基金一般项目“双向农村 O2O 商业模式的创新设计及系统整合研究”（项目编号：16BGL192）研究成果之一。
作者简介：张新香 (ORCID：0000-0003-0632-9563)，教授，博士，硕士生导师；赵彩霞 (ORCID：0000-0002-4851-0582)，数据分析师，硕士，E-mail: 1697344854@qq.com。
收稿日期：2020-06-16　　发表日期：2020-10-20　　本文责任编辑：刘远颖

响的一种方法。基于研究发现，为各影视企业进行微博营销提供实践指导。

1 文献综述

关于影响微博互动效果的探究，席林娜等[5]从情感、时间线、粉丝数等方面研究影响微博互动效果的因素，结果表明情感和粉丝数对微博的互动效果影响显著，而时间线对微博的互动效果影响不显著；陈姝等[6]结合“理性行为”原理探究微博中用户参与互动的原因，研究发现微博用户的兴趣、个人影响力、微博活跃程度等显著影响互动效果，而微博文本的表达形式对互动效果影响不明显；谢正霞[7]认为微博用户的粉丝数量正向影响互动效果，并根据这种关系得出使某一微博的互动效果达到一定程度的最佳粉丝数量；周庆山等[8]将微博中的精英用户分为：商业明星、娱乐明星和文化明星，探究他们是否对用户的互动意愿产生不同的影响，结果表明娱乐明星引起用户的互动意愿最强烈，其次是商业明星，最后是文化明星；陈娟等[9]基于政务微博数据，以转发和评论数作为微博互动效果的测量指标，探究影响辟谣微博互动效果的因素，结果发现谣言的方式和类型、微博的原创性、图片的数量、微博的互动意愿等对互动效果有显著影响；魏萌等[10]以点赞数和转发数作为互动效果的测量指标，探究“网红”微博内容的特征及其对互动效果的影响，结果表明不同类型的“网红”的受欢迎程度不同，且微博内容在互动性、丰富性、娱乐性和生动性上明显不同；S. Wu 和 J. M. Hofman 等[11]研究了 Twitter 上不同用户所发布微博的互动程度，将 Twitter 用户分为 2 个等级：一般用户与精英用户，研究表明精英用户所发布的微博更容易引起其他用户的互动；M. Cha 和 F. Benevenuto 等[12]在探究 Twitter 中信息的流行时发现已经认证过的博主所发布的文章更容易引起转发；L. Zhanglun 和 T. Q. Pengtaiquan 等[13]在探究影响微博互动效果的因素时发现文本长度对互动效果的影响显著，即较长微博文本能够促进传播的广度和速度；J. So 和 A.Prestin 等[14]以转发数作为微博的互动效果的评测指标，探究了肥胖在 Twitter 中的普遍观念和态度，结果表明贬损内容比非贬损能够更能引起用户的转发与关注，且具有幽默性的微博内容也容易引起用户的转发与关注；A. Soboleva 和 S. Burton 等[15]用转发数作为微博的互动效果的评测指标，分析了影响转发数的关键因素，结果表明账号的使用时长、好友数量以及博主的粉丝数量影响转发数，而微博文本中的标签和 URL 超链接对转发数的影响不显著。

从国内外研究现状上来看，学者们主要以微博的转发、评论等数据作为互动效果的测量指标，围绕信源、信息内容和信息接收者等方面探究微博互动效果的影响因素，而从信息内容维度出发探索微博互动效果的影响因素时，多从内容的组成形式和内容情感等方面进行探究，鲜有人探究微博内容中的隐藏主题对互动效果的影响。对于电影微博领域，更是少有研究探讨主题对微博互动效果的影响。为了弥补上述缺陷，本文尝试找出电影微博中的隐藏主题及其对互动效果影响的一种方法。基于研究发现，为各影视企业进行微博营销提供实践指导。

2 影响电影微博互动效果隐藏主题的研究方法

本文尝试结合 LDA 和多层回归模型探究电影微博中的隐藏主题及其对互动效果的影响。LDA 模型是一种常见的主题模型，能够有效识别短文本数据中的隐藏主题[16]，而多层回归模型能够分析多个层级的影响因素对互动效果的影响[17]。

2.1 LDA

LDA 是一种词袋模型，假设一篇文档由一组词组成，词和词之间没有先后关系，文档中潜在的主题将词和文档关联起来，即一个文档由一组不同概率的主题组成，一个主题由一组不同概率的词组成。在 LDA 中，文档中的词是

知识管理论坛，2020(5):283-291
DOI：10.13266/j.issn.2095-5472.2020.027

已知的，主题是未知的，根据已知的词和文本的生成规则，LDA 通过概率推测来完成参数估计。LDA 生成文本的规则如下：

抽取文档 d 中所有主题的分布关系 $\theta_d \sim Dir(\alpha)$，$Dir(\alpha)$ 是参数为 α 的 Dirichlet 分布，θ_d 是文档 - 主题的 Multinomial 分布；

抽取一个具体的主题 $z_{dn} \sim \theta_d$；

抽取主题 z_{dn} 中所有词的分布关系 $\varphi_k \sim Dir(\beta)$，$Dir(\beta)$ 是参数为 β 的 Dirichlet 分布，φ_k 是主题 - 词的 Multinomial 分布；

抽取一个具体的词 $w_{dn} \sim \varphi_k$。

重复上述步骤，不断迭代直到最终形成文档。若文档集中一共有 T 个主题，则各个主题在文档 d 中的分布概率可以表示为一个 T 维向量，这个 T 维向量满足 $\Sigma_T \theta_{d_T} = 1$。

文档中词和主题间的联合概率分布见公式（1）：

$$P(w, z|\alpha, \beta)=P(w|z, \beta)P(z|\alpha)* \int P(z|\theta)P(\theta|\alpha)d\theta \int P(w|z, \varphi)P(\varphi|\beta)d\varphi \quad 公式（1）$$

其中 w 表示文档中的词语。

2.2 回归方法

本文使用回归分析来检查隐藏主题的解释能力。一条微博接收到的转发数量、点赞数量、评论数量和用户关注数量越多，表示其引发的互动效果越强烈。在微博平台上，每条微博下面显示有微博被用户转发、点赞和评论的数量，而某一条微博吸引的新的粉丝数是不确定的，本文将每条微博的转发数量、点赞数量和评论数量作为衡量电影微博互动效果的指标。

不同电影微博的转发数量、点赞数量和评论数量存在巨大差异（见表 1），可能会引起数据的较大波动，产生高异方差性，故本文在指标处理时将参考王林等人的处理方法 [18]，利用自然对数分别将指标进行归一化，意在减小数据波动和可能存在的异方差性。最后对每一个指标进行权重赋值，则可得出每条电影微博互动效果 Y 的计算方法，如公式（2）所示：

$$Y= \alpha_1 Ln(X_1+1)+\alpha_2 Ln(X_2+1)+\alpha_3 Ln(X_3+1) \quad 公式（2）$$

其中 X_1、X_2、X_3 分别为该条微博的转发数量、评论数量和点赞数量，α_1、α_2、α_3 为其对应的权重。

王林和张玥等 [18-19] 研究认为博主的粉丝数量影响微博的转发数量，故本文将指标粉丝数量添加为自变量。话题阅读量是指在微博上以电影为中心形成的话题被微博用户访问的数量，能够表示此电影在微博用户中的欢迎程度。刘彤等 [20] 指出受众感兴趣的话题能引起更多人的关注和评论，故本文添加“话题阅读量”为自变量。

为了确定隐藏主题对电影微博互动性的影响，本文构建了因变量相同、自变量不同的两个回归模型。第一个回归模型的自变量仅包括用户粉丝数量和话题阅读量，见公式（3）；第二个回归模型添加了主题虚拟变量，即从 LDA 模型中训练得到的隐藏主题，见公式（4）。

$$Y_d^1 = \beta_0^1 + \beta_1^1 F_d + \beta_2^1 R_d + \varepsilon_d^1 \quad 公式（3）$$

$$Y_d^2 = \beta_0^2 + \beta_1^2 F_d + \beta_2^2 R_d + \sum_{k=1}^{K} \beta_k T_{k,d} + \varepsilon_d^2 \quad 公式（4）$$

其中，Y_d 表示第 d 条微博的互动效果，F_d 表示发布第 d 条微博的博主粉丝数量，R_d 表示第 d 条微博所涉及到的电影话题的阅读量。$T_{k,d}$ 为第 k 个主题在第 d 条微博中的分布概率，且 $k = 1, 2, \cdots, K$，$\Sigma_{k=1}^{K} T_{k,d} = 1$。$\beta_0$ 为常数项，β_1、β_2、β_k 为回归系数，ε_d 为随机扰动因子。

3 方法实现

3.1 数据获取

本文数据集来源于电影票房网 PC 端和新浪微博手机客户端，使用 Python 编程爬取。首先从电影票房网 PC 端获取 2017 年上映的所有电影，共 421 条，并非所有的电影能够在微博上形成以“电影名字”为主题的话题页面，故最终用于实验分析的电影有 123 个；其次以“电影名字”为搜索条件在微博上进行手动搜索，找到电影在微博上形成的话题讨论页面，选择热门微博，获取话题阅读量及每条微博的详情链接；最后根据微博详情链接获取每条微博的微博 ID、微博转发数量、微博点赞数量、微博评论数量、微博文本以及博主的用户粉丝数


知识管理论坛
2020 年 第 5 期(总第 29 期)

量，共 26 543 条，删除重复及无用微博后，最终用于实验的有 19 061 条。使用 stata 分析工具对爬取的数据进行描述统计，结果见表 1。转发、评论和点赞对应的数量最小值都为 0，最大值分别为 12 042 061、250 387、1 032 796，两者之间相差都较大(标准差明显大于均值)，且中位数明显偏向最小值的一边。用户粉丝数从 1 到 120 754 022 不等，中位数大于均值。话题阅读量的最小值为 114.3 万，最大值为 414 000 万。

表 1 变量数据的描述统计

变量	均值	标准差	最小值	中位数	最大值
转发数量(个)	2 330.21	93 216.74	0	58	12 042 061
评论数量（个）	357.40	3 339.37	0	45	250 387
点赞数量（个）	1 688.75	15 723.30	0	150	1 032 796
用户粉丝数（个）	3 445 031.21	6 767 079.49	1	688 298	120 754 022
话题阅读量（万次）	69 602.13	83 289.66	114.3	44 000	414 000

3.2 数据预处理

3.2.1 微博文本数据的预处理

LDA 模型是词袋模型，即用一组词向量表示微博文本，而原始的微博文本是由句子组成的，故在模型训练之前，使用 Python 工具对微博文本进行预处理，处理方式如下：①构建所有电影微博文本的语料库；②过滤掉无关的微博信息，包括“# 话题名 #”格式的内容、“@用户”格式的内容和文本字数小于一定值的微博文本；③使用停用词列表删除不相关和不常见的术语(例如“有”“将”“只是”)、数字和特殊字符；④把意思相近的词语合并成同一个术语，例如“笑点”和“笑料”被组合成“笑点”；⑤不同词性的词对主题表达的贡献程度不同，对主题表达和辨识作用最大的是名词和动词 [21]，故本文只考虑名词和动词作为特征词；⑥分析发现词频具有明显的“长尾特征”，即多数词出现较小的次数，少数词出现较高的次数，这和唐晓波等 [22] 的结论一致。本文使用 0.99 级别的 RemoveSparseTerms 函数，删除所有文档中出现小于 1% 的动词和名词 [23]，这是一种常见的做法，允许研究人员获得更短的术语列表和更多有用的信息。经过预处理，最终用于 LDA 训练的词语有 393 个，使用 Python 的 wordcloud 库绘制了电影微博文本词频的词云图，如图 1 所示：



图 1 电影微博文本词频的词云

3.2.2 变量的预处理

使用 Python 工具对变量进行预处理，处理方式如下：将所有变量进行归一化，第 i 条微博的第 j 个变量表示为 x_{ij}，标准化后的 $x'_{ij}=\frac{x_{ij}-\min(x_j)}{\max(x_j)-\min(x_j)}$，其中 $min(x_j)$ 表示变量 x_j 的最小值，$max(x_j)$ 表示变量 x_j 的最大值，j=1, 2…, 5, i=1, 2…, n；

本文使用信息熵计算转发、评论和点赞量 3 个指标的权重 [17]。信息熵是刻画信息无序度的一个指标，熵值越大，表示信息的无序化程度越高，相对应的信息效率越高 [24]。计算变量转

知识管理论坛，2020(5):283−291
DOI：10.13266/j.issn.2095−5472.2020.027

发数量、评论数量和点赞数量的信息熵：$e_j = -\ln(n)^{-1}\sum_{i=1}^{n} p_{ij}\ln p_{ij}$，$p_{ij} = \frac{x'_{ij}}{\sum_{i=1}^{n} x'_{ij}}$，得出 e_1=0.992 2、e_2=0.994 2、e_3=0.989 7。

计算各变量的权重 $\alpha_j = \frac{1-e_j}{\sum_j 1-e_j}$，结果得出 α_1=33%、α_2=43%、α_3=24%。

根据公式（2）计算出每条微博的互动效果，将其与归一化后的转发数量、评论数量和点赞数量这 3 个指标分别做相关分析，发现相关系数分别为 0.91、0.9、0.94，均大于 0.9，说明权重赋值合理[17]。

3.3 试验结果

3.3.1 LDA 结果

使用 Python 中的 scikit-learnLDA 包训练电影微博文本数据。LDA 最重要的组成部分之一是确定构成所有文本的主题数量，本文使用 perplexity 指标[25]来确定微博文本的最佳主题数量，测试主题数量编号从 2 到 15，结果表明最佳选择是 6 个主题，见图 2。使用 pyLDAvis 包来可视化主题映射，图 3 显示 6 个主题都在自己的非重叠区域中，这表明 6 个主题之间没有相关性，每个主题都是独立且信息丰富的。

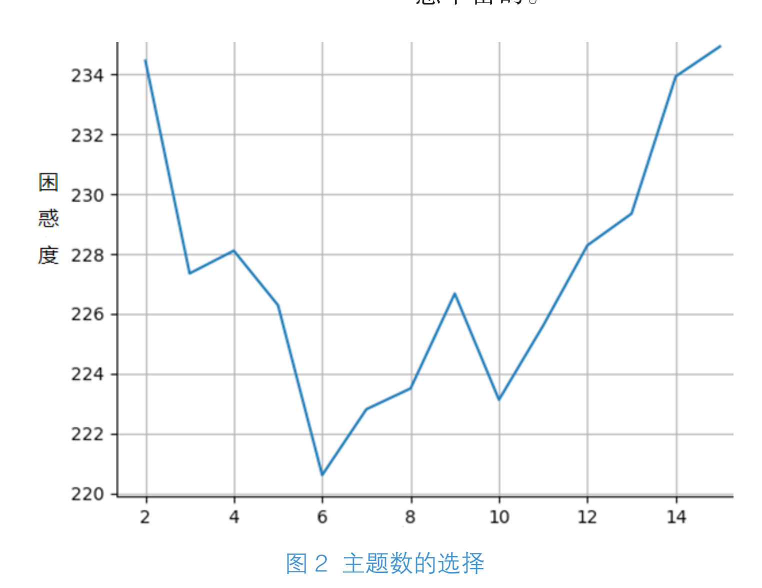


图 2 主题数的选择



图 3 6 个主题的可视化映射图



LDA 的结果中每个主题由一组词组成。一个好的主题模型不仅取决于模型在可衡量的统计指标中的表现，还取决于每个主题的合理性和可解释性。每个主题中的前 15 个主题词见表 2。LDA 允许一些词语出现在多个主题中，例如，词语“导演”出现在主题 1 和主题 6 中，词语“发布”出现在主题 2 和主题 4 中。表 2 中的结果显示主题 1 包含相关电影人物介绍方面的术语，例如“导演”“演员”“角色”“饰演”，因此本文将主题 1 命名为“电影人物”；主题 2 的主题词更可能描述了有关电影宣传的微博，例如“预告”“海报”“预告片”“公映”，因此本文将主题 2 命名为“电影宣传”；主题 3 包含互动营销方面的术语，例如“转发”“关注”“机会”“赠送”，因此本文将主题 3 命名为“互动营销”；主题 4 更可能是有关电影主题、情节、插曲的微博，因此本文将主题 4 命名为“电影内容”；主题 5 主要包含电影观后感及评价方面的术语，例如“喜欢”“支持”“好看”“搞笑”，因此本文将主题 5 命名为“电影评价”；主题 6 包含电影线下宣传活动方面的术语，例如“现场”“发布会”“路演”“首映礼”，因此本文将主题命名为“线下活动”。审查文本产生的所有 6 个不同主题都是可解释和合理的。

表 2 每个主题中的前 15 个主题词

主题1（电影人物）	主题2(电影宣传)	主题3(互动营销)	主题4(电影内容)	主题5(电影评价)	主题6(线下活动)
导演	预告	转发	故事	喜欢	现场
演员	版本	微博	爱情	感谢	发布会
角色	发布	关注	青春	开心	导演
饰演	海报	电影票	人类	支持	直播
拍摄	饰演	机会	主题曲	朋友	首映
影片	主演	评价	演唱	搞笑	采访
片段	曝光	抽取	世界	可爱	宣传
故事	执导	观影	音乐	笑点	主创
幕后	档期	赠送	热血	祝福	路演
演技	预告片	好友	人生	推荐	首映礼
主角	期待	电影院	歌曲	观众	视频
特效	特辑	评论	发布	剧组	活动
动作	影片	全国	版本	彩蛋	主演
镜头	公映	热映	讲述	大卖	粉丝
制作	监制	影片	上线	热映	亮相

3.3.2 回归结果

使用 stata 分析工具对数据进行回归处理，结果见表 3。模型 1 中，所有变量的系数在 1% 水平上具有显著性，“用户粉丝数”和“话题阅读量”的系数均为正，这表明“用户粉丝数”和“话题阅读量”正向影响微博的互动效果。在模型 2 中，变量“用户粉丝数”及“话题阅读量”的系数和模型 1 中的结果基本相当。除了模型 1 中的自变量之外，模型 2 还添加了 LDA 中的隐藏变量。由于自变量矩阵 X 中的项 $T_{k,d}$ 满足 $\sum_{k=1}^{K} T_{k,d}=1$，为了避免自变量间的多重共线性的产生，本文参考 H. Yan 等 [25]

知识管理论坛，2020(5):283-291
DOI：10.13266/j.issn.2095-5472.2020.027

的做法，将主题 1 从模型中删除。对主题 1 之外的所有自变量进行多重共线性检验，结果见表 4，方差膨胀因子 VIF 的 *Mean*=1.44>1，*Max*=1.75<10，故变量之间不存在多重共线性。模型 2 结果显示主题 2、主题 3、主题 4 和主题 5 在 1% 水平上都具有统计显著性，且系数均为正，表明电影微博文本中有关“电影宣传”“互动营销”“电影内容”和“电影评价”的主题正向影响微博的互动效果，可能对微博的互动效果有促进作用。

表 3 回归结果

变量	模型1				模型2			
	Coef.	Std. Err.	z	P>|z|	Coef.	Std. Err.	z	P>|z|
用户粉丝数	0.221 9	0.003 7	60.32	0.000	0.234 1	0.003 8	61.61	0.000
话题阅读量	0.236 9	0.009 1	26.05	0.000	0.236 9	0.009 0	26.27	0.000
主题2（电影宣传）					0.171 8	0.054 5	3.15	0.002
主题3（互动营销）					1.131 4	0.057 5	19.67	0.000
主题4（电影内容）					0.187 4	0.057 8	3.24	0.001
主题5（电影评价）					0.757 3	0.058 6	12.93	0.000
主题6（线下活动）					0.093 2	0.059 2	1.58	0.115

表 4 多重共线性检验

主题变量	VIF	1/VIF
主题2	1.75	0.570 259
主题3	1.53	0.654 314
主题4	1.53	0.651 727
主题5	1.63	0.613 202
主题6	1.52	0.656 697
用户粉丝数	1.1	0.906 167
话题阅读量	1.02	0.983 144
平均 VIF	1.44	

4 总结与讨论

4.1 总结

本文通过结合 LDA 和 2 层回归模型的方法找出了电影微博中的隐藏主题及其对互动效果的影响。本文研究主要结论如下：

（1）使用主题挖掘模型 LDA 发现了隐藏在电影微博文本中的 6 个可解释主题：“电影人物”“电影宣传”“互动营销”“电影内容”“电影评价”和“线下活动”，主题可视化映射结果表明 6 个主题之间没有相关性，每个主题都是独特且信息丰富的。

（2）构建 2 层回归模型分析 6 个隐藏主题对电影微博互动效果的影响。结果表明变量“用户粉丝数”和“话题阅读量”正向影响电影微博互动效果。添加主题虚拟变量后增加了模型的解释能力，主题“电影宣传”“互动营销”“电影内容”以及“电影评价”对微博互动效果影响具有显著性。

4.2 营销建议

本研究结果能够为企业开展电影微博营销提供一些启示和建议。

首先，回归结果显示“用户粉丝数”和“话题阅读量”对电影微博互动效果影响显著。微博大 V 往往具有庞大的粉丝数量，企业可以通过与微博大 V 合作，使电影得到更广泛的互动和宣传。

其次，回归结果同时表明主题“电影宣传”“互动营销”“电影内容”和“电影评价”正向影响电影微博互动效果。企业一方面可以在官方微博中发布一些有奖转发之类的活动以及和电影有关的歌曲、海报、主题、情节、精

彩片段；另一方面，企业应主动邀请一些在微博中有影响力的人物观看电影并对电影做出相关评价。

4.3 研究不足与展望

本研究有一些限制，希望未来能进一步完善这些不足。首先在用 perplexity 指标选择最佳主题时，只测试了主题数为 2-15 的模型困惑度，对主题数大于 15 的情况没有考虑。未来研究可以测试更多主题数，挖掘隐藏在电影微博文本中更细粒度的主题。其次主题的可解释性来自主观判断，未来将采用专家调研和内容分析等技术科学获取主题的命名，以提升主题的可解释性。

参考文献：

[1] 夏立新, 张纯, 陈健瑶, 等. 企业微博内容对网络口碑及品牌认可度的影响 [J]. 情报科学, 2019, 37(4):79-85.

[2] 李慧, 王丽婷. 基于话题标签的微博热点话题演化研究 [J]. 情报科学, 2019, 37(1):30-36.

[3] 新浪微博数据中心 .2018 年度微博电影白皮书 [EB/OL].[2019-02-03].https://data.weibo.com/report/reportDetail?id=430.

[4] 李红, 项子晴. 企业微博营销效果影响因素分析——以小米公司为例 [J]. 价格月刊, 2017(11):78-80.

[5] 席林娜, 窦永香. 基于计划行为理论的微博用户转发行为影响因素研究 [J]. 数据分析与知识发现, 2019, 3(2):13-20.

[6] 陈姝, 窦永香, 张青杰. 基于理性行为理论的微博用户转发行为影响因素研究 [J]. 情报杂志, 2017, 36(11):147-152, 160.

[7] 谢正霞. 基于用户交互的企业微博传播效果评价研究 [D]. 广州 : 华南理工大学, 2019.

[8] 周庆山, 赵雪, 赵需要, 等. 我国数字内容产业研究的主题知识图谱分析 [J]. 情报理论与实践, 2012, 35(4):56-61.

[9] 陈娟, 刘燕平, 邓胜利. 政务微博辟谣信息传播效果的影响因素研究 [J]. 情报科学, 2018, 36(1):91-95, 117.

[10] 魏萌, 张博. 新浪微博"网红"的微博内容特征及传播效果研究 [J]. 情报科学, 2018, 36(2):88-94.

[11] WU S, HOFMAN J M, MASON W A, et al. Who says what to whom on twitter[C]//Proceedings of the 20th international conference on world wide Web. Hyderabad: ACM, 2011: 705-714.

[12] CHA M, BENEVENUTO F, HADDADI H, et al. The world of connections and information flow in twitter[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part A: systems and humans, 2012, 42(4): 991-998.

[13] ZHANG L, PENG T Q. Breadth, depth, and speed: diffusion of advertising messages on microblogging sites[J]. Internet research, 2015, 25(3): 453-470.

[14] SO J, PRESTIN A, LEE L, et al. What do people like to "share" about obesity? a content analysis of frequent retweets about obesity on Twitter[J]. Health communication, 2016, 31(2): 193-206.

[15] SOBOLEVA A, BURTON S, MALLIK G, et al. 'Retweet for a Chance to…': an analysis of what triggers consumers to engage in seeded eWOM on Twitter[J]. Journal of marketing management, 2017, 33(13/14): 1120-1148.

[16] 周炜翔, 张仰森, 张良. 面向微博热点事件的话题检测及表述方法研究 [J]. 计算机应用研究, 2019, 36(12):3565-3569, 3578.

[17] 刘艳秋, 方建群, 胡静, 等. 人格因素及家庭环境对生态移民青少年冒险行为的影响 [J]. 中国健康心理学杂志, 2019, 27(12):1871-1875.

[18] 王林, 潘陈益, 朱文静, 等. 机构微博传播力影响因素研究 [J]. 现代情报, 2018, 38(4): 35-41.

[19] 张玥, 孙霄凌, 浦正宁, 等. 微博舆情传播影响因素研究——基于信源特征和信息形式的视角 [J]. 情报资料工作, 2014(3): 59-64.

[20] 刘彤, 尹仔锋. "可及"环境与"不可及"环境——电影微博营销影响因素探析 [J]. 新闻界, 2012(23): 66-70.

[21] 李莉, 林雨蓝, 姚瑞波. 基于 LDA 模型的交互式文本主题挖掘研究——以客服聊天记录为例 [J]. 情报科学, 2018, 36(10): 64-70.

[22] 唐晓波, 王洪艳. 基于潜在语义分析的微博主题挖掘模型研究 [J]. 情报学报, 2013, 56(24): 114-119.

[23] HENG Y, GAO Z, JIANG Y, et al. Exploring hidden factors behind online food shopping from Amazon reviews: a topic mining approach[J]. Journal of retailing and consumer services, 2018 (42): 161-168.

[24] 金坚. 大数据时代信息熵的价值评价研究 [D]. 长春 : 吉林大学, 2019.

[25] WANG Y, AGICHTEIN E, BENZI M. TM-LDA: efficient online modeling of latent topic transitions in social media[C]// KDD '12: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining. New York: Association for Computing Machinery, 2012: 123–131.

作者贡献说明：

张新香：指导论文构思与写作，提出修改意见并修改终稿；

赵彩霞：负责数据采集、初稿撰写及论文修改。

知识管理论坛，2020(5):283–291
DOI：10.13266/j.issn.2095–5472.2020.027

Research Method and Application of Hidden Themes Influencing the Interactive Effect of Movie Microblog

Zhang Xinxiang Zhao Caixia
School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073

Abstract: [Purpose/significance] Exploring the hidden themes that affect the interactive effect of movie microblogging can explore the hot issues of users' attention and provide effective marketing strategies for enterprises. **[Method/process]** This paper crawled the popular microblog of 123 movies released in 2017 from Sina Weibo, used the topic modeling method to mine the hidden themes in the movie microblog text, and used the regression method to analyze the impact of hidden themes on the interactive effect of movie microblogging. **[Results/conclusions]** It turns out that there are 6 interpretable themes: movie characters, movie promotion, interactive marketing, movie content, movie evaluation and offline activities, of which 4 themes of movie promotion, interactive marketing, movie content and movie evaluation have a positive impact on the interactive effect of movie Weibo; at the same time, it is found that the number of user fans and the popularity of topic discussion positively affect the interactive effect of movie Weibo.

Keywords: movie microblog   interactive effect   topic model   LDA